DOCUMENT RESUME

ED 351 018 IR 054 212

TITLE	Data Conversion and Indexing: How To Get the Most out of Your Machine-Readable Database. DataResearch Automation Guide Series, Number Two.
INSTITUTION	Data Research Associates, Inc., St. Louis, MO.
PUB DATE	Jan 92
NOTE	11p.; For other reports in this series, see IR 054 211-213.
PUB TYPE	Guides - Non-Classroom Use (055) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Automatic Indexing; *Bibliographic Utilities; *Data Processing; Indexing; *Library Automation; *Machine Readable Cataloging; *Online Catalogs; Services; Subject Index Terms
IDENTIFIERS	*Authority Control (Information); *Data Conversion; MARC; Retrospective Conversion (Library Catalogs)

ABSTRACT

In today's library environment, the library's most important asset is its machine-readable database. In order to maintain this investment, Data Research Associates provides several comprehensive data conversion services, including retrospective conversion and authority control. For bibliographic records not yet in MARC (machine readable cataloging) format, Data Research offers network cataloging and batch processing services. Network cataloging allows the library to connect to the complete LCMARC database and overlay records manually. If the records contain a numeric search key they can be extracted and run against the LCMARC database for matching. Authority records allow the library to improve the accuracy and consistency of headings used in the bibliographic records. Authority records, most of which have been created and distributed by the Library of Congress, are available for names, uniform titles, series, and subjects. Data Research will take the library's bibliographic records and authorize the database, which will eliminate the variations found within headings that are identical from a content perspective but considered different by the computer. Once the library's database has been authority processed, Data Research can provide online authority verification to keep its file current. To assist with constructing an index, Data Research can also provide a Standard Indexing Table for Bibliographic and Authority Records. (KRN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED351018

Data Conversion and Indexing

How to Get the Most out of Your Machine-Readable Database

SINE MURIS LIBRARY

3 9349 00376 0473

120542/2

DATA RESEARCH

AUTOMATION GUIDE SERIES

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Carl Grant

TO THE EDUCATIONAL RESOURCES

DATA CONVERSION SERVICES

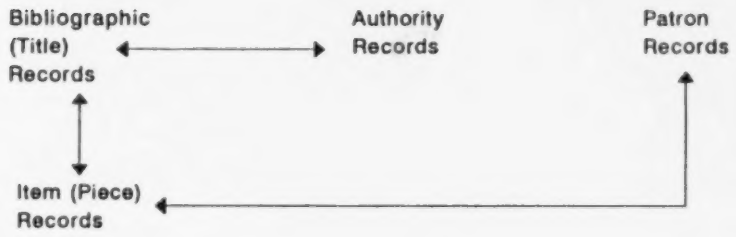
Overview

In today's ever-changing library environment, the most important asset a library has is its machine-readable database. Since resources are rarely available for anything other than routine maintenance and updating, this database will only be created once. Thus, it will outlive the library's current automated system, as well as future replacement systems. In order to maintain this investment, Data Research provides several comprehensive data conversion services, including retrospective conversion and authority control.

Although an automated library system has many files that contain the library's data, four are of particular interest, including:

- Bibliographic (title) records
- Authority records
- Item (volume) records
- Patron (user) records

Figure 1: Relationship of Major Files



Bibliographic Records

MARC Records

Within the library community there is a standard for communicating these machine-readable records known as MARC (MAchine-Readable Cataloging), which is defined by the National Information Standards Organization (NISO) in its Z39.2 standard. The attractive features of the MARC record are that it is designed to accommodate the needs of variable length fields and records and that some fields can be repeated as needed.

Source of MARC Records

Sources of MARC bibliographic records include:

- Library Automation Vendors, such as Data Research
- Library of Congress
- Bibliographic Utilities, including OCLC, RLIN, WLN and UTLAS
- Book Jobbers, including Baker & Taylor, Brodart and Ingram
- COM/CD-ROM Vendors, including Library Corporation, General Research Corporation, Library Systems & Services, Inc. and Marcive

Records may be provided using an online system or distributed via a CD-ROM or floppy diskette. In most cases, records may also be transferred or downloaded from one system to another.

Although pricing is an important concern, libraries should also take into account the percentage of their collection that will be found in the database source they are considering. If the percentage or hit rate is relatively low, then the library may need to consult a second source. For those bibliographic records not found, typically the library will need to perform (or pay to have done) original cataloging, in order to create the necessary machine-readable bibliographic record.

Before purchasing a local library automation system, the library should verify whether the system will allow the library to enter, edit, store and output a complete MARC record -- including the leader. While this seems like a rather obvious requirement, libraries must be aware since some systems do not comply.

Existing Record Machine Conversion

Many libraries will already be using a utility for cataloging. These records are readily converted for loading on an automated system. The major issues in these cases will be indexing, duplication and filing indicator checks.

Retrospective Conversion

Frequently the source used for retro-conversion can also be used for ongoing catalogs. For those bib records not yet in MARC format, the library must perform retrospective conversion.

For these sites Data Research offers network cataloging and batch processing services.

Network Cataloging

Network cataloging, a service unique to Data Research, allows the library to connect to the complete LCMARC database housed on the Data Research headquarters computer and overlay records manually. Authority and bibliographic services are available.

Batch Processing

If the records contain a numeric search key, such as an OCLC number, LC card number or ISBN these can be extracted and run against the LCMARC database for matching. Records not matched via this method can be manually upgraded by mapping the non-matched records to MARC tags in the form of brief records. These can be loaded into the library's system and are then available for circulation, the public access catalog, etc. and can be cleaned up as time permits. In the event that the library decides to proceed with upgrade services in the future, Data Research provides a print-out of the non-matched records.

Conversion for Sites With No Machine-Readable Data

For those sites that do not have machine-readable data at all, Data Research offers pre-packaged databases or manual-entering of search keys by library staff. Alternatively, we can provide for the library manual contract service for manual entering of search keys and MARC tagged records.

Pre-packaged Databases

Based on parameters concerning year of publication, bibliographic type and reading level, a set of records can be pulled from the LCMARC database for use by the library. The library can then attach items to the records and at a specified date, delete those records with no holdings attached. The pre-packaged databases can also include authorities.

Manual Entering of Search Keys by Library Staff

The site can manually enter search keys, such as the OCLC number, LC Card number and/or ISBN using programs available from Data Research. These keys are then used as input for the process described in the batch processing section above.

Contract Services for Manual Entering

Data Research can subcontract for the keying of search keys and then will run the resulting file against LCMARC as described above.

Those records that cannot be matched can be keyed to MARC specifications. The authorized headings and forms of names will be keyed as they appear on the card numerically, i.e., LCCN, ISBN. In addition to MARC tapes, access to OCLC tapes on search keys via OCLC is available. These services apply to the actual shelflist, photocopies of the shelflist, microfilm, copyflo hardcopy, and abbreviated records on diskette or tape. It is also possible to create search keys if there is enough information in magnetic form. With respect to holdings in all of these service levels, it is possible to key holdings if the library opts for it. When the search string is keyed, the 949 field is added, which includes a holdings symbol and other holdings data. If it is discernible that there are multiple copies on the card, multiple 949 fields will be entered.

Authority Records

Value of Authority Records

With the increasing trend toward shared resources, regional networks and merged databases, authority records play a vital role in the initiation and maintenance of bibliographic databases. Authority records allow the library to improve on the accuracy and consistency of headings used in the bibliographic records, since name variations can be traced, nicknames can be identified, and related subjects can be linked together. Moreover, the authority record can be used to direct the user of the library's catalog to the authorized heading (SEE cross-reference) or entry and suggest other headings that might be of interest (SEE ALSO cross-references). Several research studies suggest that there is a direct correlation between the number of cross-references and the search success rate experienced by the user of the catalog. Authority records are available for names (personal, corporate, and conference), uniform titles, series and subjects. The majority of authority records have been created and distributed by the Library of Congress.

Typically, it is better for a library to have authority processing completed prior to loading and indexing on the local system. If this is done, both staff and patrons will see a clean database at the beginning of use rather than at some point in the future.

Authorization of Database

Rather than create their own authority records for every heading found in a bibliographic record, most libraries employ the services of a vendor, like Data Research, who will take the library's bibliographic records and "authorize the database." This database clean-up will eliminate the variations found within two identical headings from a content perspective that are considered different by the computer. For example, consider the headings:

User interface (Computer systems)
User interface (Computer systems)
User interface (computer systems)
User interface--Computer systems
User interface-Computer systems

Each heading is identical from a content point-of-view, but would be treated differently by the computer due to variations in capitalization, punctuation and spacing.

Note: The presence of these four headings in a library's catalog would probably thoroughly confuse the average patron. Using the services of Data Research, these four headings would be combined and the "authorized" heading, **User interface (Computer systems)**, would be inserted in each bibliographic record as necessary.

In some cases, the computer itself cannot be programmed to discover and automatically correct these variations in headings or in headings for which no SEE or SEE ALSO cross-reference can be found. Usually in these cases, a knowledgeable trained cataloging librarian is used to make the necessary professional judgements about what heading to use and whether to add a cross-reference to an authority record. This authorization process also cleans up typographical errors.

Once complete, the authorized bibliographic records are returned to the library. These authorized bibliographic records can then be overlaid onto the corresponding bibliographic record that is already loaded into the library's automated library system.

Loading Authority Records

At the option of the library, the vendor can also provide the corresponding authority records that were used by the bibliographic records. These machine-readable authority records can then be loaded into the library's automated library system. This allows the library to easily update these authority records with additional cross-references in order to make the library's collection more accessible to their patrons.

In most cases, the library can also create its own local authority control records, complete with SEE and SEE ALSO cross-references. Some libraries have even purposely added "typos" as a cross-reference, e.g., spelling variations for "Tchaikovsky"!

Online Authority Verification

Once the library's database has been authority processed, it must select a means for keeping this file current. The library can schedule, at specific intervals, another pass of its database against the authority files. If it is a Data Research library, it can subscribe to Data Research's Online Authority Verification Service. This service allows the library to verify new bibliographic records against the current LC files as they are added to the library's database and to quickly obtain updates to existing authority records. The library is immediately notified of any non-conformance so that immediate action may be taken to correct, add or delete the flagged field. If a new authority record exists in the LC files, it may be copied to library's database instantly. Overall, this service offers you a method of keeping your investment in authority control current and increases access to materials for your users.

Item Records

Contents of Item Records

Item records contain the item (or volume) specific information, e.g., branch location, collection location, call number, part number, barcode number, status (on-the-shelf, checked out, lost), etc. These item records are linked to the appropriate bibliographic (title) record. The amount of data stored in the item record is important. A longer or larger item record will allow the library to prepare reports detailing the usage of the collection based on a variety of selection criteria.

Patron Records

Creating Patron Records

Patron records can be created by manually entering the data into the automated library system using a patron registration screen. Libraries that have made their paper registration form look similar to the online registration screen have found that the data entry process goes much faster and results in fewer errors being made.

Typically, libraries will order a pair of barcode labels or an additional eye-readable strip with the same barcode number. The barcode label is affixed to the library patron registration card and the additional strip containing the number is affixed to the paper patron registration form to minimize number transposition errors.

In some cases, usually in academic libraries, it's possible to use the Registrar's file to create the necessary patron records. A tape, conforming to a specified record format, is provided by the Registrar or the vendor can provide the necessary custom programming to create the patron records.

The amount of information stored in the patron record is important. With a larger patron record, the library is able to prepare a detailed analysis of usage of the library's collection by different patron groups.

Indexing

Indexing Options

With most automated library systems, the library has the option of specifying what fields and subfields are used to construct each index. In addition, the library is able to specify what data consistency checks the computer makes when loading and indexing each record. The library is normally able to specify what indexes are created for the bibliographic and authority records. Indexes for the item and patron files are predetermined by the vendor.

To assist in that process Data Research will provide a Standard Indexing Table for Bibliographic and Authority Records. A Data Research Installation Specialist is available to answer any questions that might arise concerning indexing.

Data Research can create an index for a library's database using computers located in St. Louis, headquarters for Data Research. During the time the library is waiting for its own computer system to be delivered and installed, the library sends its database to St. Louis for processing.

Once the library's computer has been installed, the library's database, complete with associated indexes, is loaded (restored) into the library's computer. This latter process takes only a matter of a few hours. Thus, the library can begin training on its new automated library system using its own database.

The library also has the option to load and index its database after its computer has arrived and been installed. The speed of indexing will vary depending upon the size of the databases, the processing power of the computer, the number of indexes to be created, and the "intelligence" of the software. Be sure to check with other customer libraries of your preferred vendor to determine the speed with which your library's database can be loaded and indexed. Some vendors can take several months to achieve this task.

The Guide Series is designed to provide concise, practical and relevant information about a variety of topics related to automated library systems. Topics to be addressed in this series include:

The following Guides AVAILABLE at this time:

1 Opening Up "Open Systems" Guide
2 Data Conversion and Indexing Guide
3 Successfully Automating Library Consortia

The following Guides are NOT YET available:

4 Guide to Selecting an Automated Library System
5 Library Automation Benefits Guide
6 Retrospective Conversion Guide
7 Barcoding Guide
8 Computer Site Preparation Guide
9 Power Conditioning Guide
10 Installation Guide
11 Telecommunications Guide
12 Guide to Local Area Networks
13 Guide to Managing an Automated Library System
14 Mobile Library Automation Guide
15 System Migration Guide

When ordering please specific the number of the Guide(s) you wish to receive. Guides concerning other topics may be added from time to time. Up to three Guides are provided compliments of Data Research. Copies of additional Guides may be ordered from Data Research for $3.00. Comments on the Series or a particular Guide are welcomed.

Please contact:

Marketing Department
Data Research Associates, Inc.
1276 North Warson Road
St. Louis, Missouri 63132-1806

800-325-0888 314-432-1100 314-993-8927 (fax)

NOTES

Copyright© 1992. Data Research Associates, Inc. All Rights Reserved.

Printed on recycled paper.

Data Research Associates, Inc.
(800) 325–0888
(314) 432–1100
(314) 993–8927 fax
Internet: Sales@DRANET.DRA.COM